

UNITED STATES UTILITY PATENT APPLICATION

for

POWER MANAGEMENT STRATEGY TO SUPPORT
HOT SWAPPING OF SYSTEM BLADES DURING RUN TIME

Inventor:

Hong W. Wong

Attorney Docket No.: 042390.P9705

Prepared by:

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN, LLP
12400 Wilshire Boulevard, Seventh FL
Los Angeles, California 90025-1026
408-720-8300

EXPRESS MAIL CERTIFICATE OF MAILING

"Express Mail" mailing label number EL627467067US Date of Deposit September 29, 2000
I hereby certify that I am causing this paper or fee to be deposited with the United States Postal Service
"Express Mail Post Office to Addressee" service on the date indicated above and that this paper or fee
has been addressed to:

Box Patent Application
Commissioner of Patents
& Trademarks
Washington, D. C. 20231


by Julie K. Mausen
Julie K. Mausen

Date Signed: September 29, 2000

POWER MANAGEMENT STRATEGY TO SUPPORT
HOT SWAPPING OF SYSTEM BLADES DURING RUN TIME

BACKGROUND OF THE INVENTION

5 1. Field of the Invention.

The invention relates to the field of systems with hot swappable components.

10 2. Prior Art and Related Art.

In many applications it is necessary to have a system which remains in operation even when under repair. For instance, a server in a network may have boards or blades in a standby mode or with excess capacity. When a blade is detected as having failed, other blades that were in a standby mode take over while the failed blade is replaced. Alternatively, blades with excess capacity are used to pick-up the tasks of the failed blade while it is replaced. Generally, an alarm or other indication is given to an operator indicating that a failure has occurred. The operator is then able to remove the failed blade and reinsert a good blade. One of the major problems in these systems is the management of power as these changes occur.

Often in such systems connectors are used having pins of different lengths which cooperatively engage sockets. For instance, a backplane in a server may have a plurality of connectors, each for receiving a blade where the pins in the connectors are of different lengths. These pins are used for

instance, to allow discharge of static electricity as well as other functions.

00000000000000000000000000000000

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a block diagram of a system built in accordance with the present invention.

Figure 2A is a flow diagram illustrating the events occurring during the 5 normal insertion of a blade into the system of Figure 1.

Figure 2B is a flow diagram illustrating the events occurring during the normal extraction of a blade from the system of Figure 1.

Figure 3 is a flow diagram illustrating the events occurring during the forced extraction of a blade from the system of Figure 1.

10 Figure 4 is a flow diagram illustrating the events occurring during the failure process of a blade in the system of Figure 1.

Figure 5 is a diagram illustrating a backplane, a connector with pins of three different lengths and a blade which cooperatively engages the connector in accordance with the present invention.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

A system is described having a backplane bus and a plurality of blades or boards which may be extracted or inserted while the system is in operation. This is often referred to as "hot swapping."

5 In the following description, numerous specific details are set forth, such as specific voltages in order to provide a thorough understanding of the present invention. It will be apparent to one skilled in the art that the present invention may be practiced without these specific details. In other instances, well-known circuits and other components have not been set forth in detail in
10 order not to obscure the present invention.

The present invention is described in connection with a server having hot swappable blades. Referring to Figure 1, the system includes a backplane bus only a portion of which is shown in Figure 1. Lines 13 and buses 30 are the only portion of the backplane shown, power and ground lines are not shown. The buses shown by the bracket 30 include the Interchassis Management Bus (IMB), Universal Serial Bus (USB), Keyboard/Video/Mouse (KVM) bus, I²C bus, RS232 port and a fan control bus. Other buses may be used. While the system management bus (SM Bus) is shown as part of the backplane in Figure 1, the SM Bus in one
15 embodiment is a separate bus interconnecting the blades in the system.
20

In one embodiment, the server includes several different units which may be hot swapped and which engage the backplane bus. These include the two server management blades (MBs) shown in Figure 1 as blades 10 and 12. The server also includes at least one switch blade 14 and typically 5 include 2 more switch blades. These switch blades provide network switching of packets and routing for various protocol layers such as layers 2, 3, 4 to 7. Additionally there are a plurality of central processing unit (CPU) blades, for instance, 38 CPU blades may be used or even a higher number. Two such blades 15 and 16 are shown in Figure 1. These blades perform 10 the normal computing functions required of a server.

While not shown in Figure 1, the system also includes at least one power sharing board for distributing power (48V DC in one embodiment) to the various blades in the system from the backplane. Also, multiple fans are used for cooling which is often the case in such systems.

15 The present invention provides operating characteristics for the server which meet the predetermined criteria shown below in a table.

TOU70

Hot Swap States	Blade Status				CPU Blade	Switch Blade	System Mgmt Blade
	Operation State	3V or 5V stby	Reset	48V DC Cvr			
P0	The blade is physically separated from the system				Yes	Yes	Yes
P1/H0	The blade is fully seated	On	On	Off	Yes	Yes	Yes
H1	The blade is fully seated and powered up	On	On	On			
H2	The blade is allowed to boot up to BIOS/Embedded SW/Diagnostics partition	On	Off	On	Yes	Yes	Yes
H1F	The blade is allowed to boot up, but the board has detected an error and powered itself down. BIOS/embedded SW stopped running	On	Off	Off, over-ridden by blade	Yes	Yes	Yes
H3	The blade is powered up and is operating in its normal state, with OS or Embedded SW running	On	Off	On	Yes	Yes	Yes

In the table above the hot swap states include P0 which is the state where the board is removed from the backplane and P1, the state where a blade is fully inserted into the backplane. The states H0 through H3 indicate several operational states for the blade as indicated in the above table. The "Yes" in the last three columns of the table indicate that the CPU blade, switch blades and SMBs support the indicated hot swap states.

While the server operates primarily on 48V DC there is also a 3V or 10 5V standby power used for a sleep mode. The state of that power is shown under "5V stby" in the above table. The system also employs a reset signal which is discussed below and shown in the table. The power management

apparatus of the present invention, as will become apparent from the description below, provides the capability shown in the table.

Each of the blades includes a power supply which, in one embodiment, is a DC-to-DC converter for filtering and regulating the 48V power. The converter is enabled with an enable signal as will be discussed. The state of the converter is shown as either "off" or "on" in the above table.

One of the major problems in a hot swappable system is providing power to the blades and maintaining the power when the blades are hot swapped. The removal and insertion of blades often causes transients in the power delivery system which affect the operation of other blades and the power subsystem.

In one embodiment of the present invention each of the blades includes a female connector and a corresponding male connector in the backplane having pins of three different lengths. This is shown in Figure 5 where backplane 100 is shown along with a male connector 101. The male connector 101 has a plurality of pins only a few of which are shown in Figure 5. Each of the pins are one of three lengths, a first and longest length such as pin 106; the shortest length such as pin 102; and, a medium length pin which is longer than the shortest length pin and shorter than the longest length pin such as pin 103. The corresponding female socket in each of the

blades such as blade 108 of Figure 5 includes a plurality of sockets each of which cooperatively engages its respective pin.

While typically the male connector is part of the backplane and the female connector is on the blade, in another embodiment the male connector

5 may be on the blade and the female connector in the backplane. In this patent "pin/socket" is sometimes used to indicate that either the pin or the socket is being designated, depending on whether the backplane has the male or female connector. The term pin/socket indicates, for instance, that the socket cooperatively receives a pin of a designated length.

10 As is apparent from Figure 5, as the blade 108 is inserted, electrical contact will first be made between the longest pins such as pin 106 and corresponding sockets such as socket 121. As insertion continues, the medium length pins next make contact such as the pin 103. And finally, as insertion continues the shortest pins such as pins 102 and 105 make contact

15 with sockets 120 and 125, respectively. Likewise on extraction, first the shortest pins/sockets part, followed by the medium pins/sockets, and finally the longest pins/sockets separate.

Each of the blades includes a power supply which in one embodiment is a 48V DC-to-DC converter. As shown in Figure 5 the converter 109 has

20 an input terminal which receives input power for the blade from lines in the backplane 110. Each of the converters has an output terminal 111 which

provides regulated DC power for the blade. The converters also has an enable terminal 112 that controls an enable function in the converter. This turns the converter "on" or "off". When turned off, that is when disabled, the input terminal 110 is effectively decoupled from the output terminal 111.

5 This decouples the substantial capacitance typically found at the output of a converter from terminal 110.

The input terminal to the converter receives power from two different lines. A first line 114 is connected to the shortest length of the three pin/socket lengths (pin 115 and socket 122). The input terminal 110 also

10 receives power from a long pin/socket namely pin 106 and socket 121. This pin/socket is connected to the input terminal 110 through a resistor 116. This resistor as will be discussed, plays an important role in dampening transient currents to the blade. The ground connection to the blade is through a long pin/socket.

15 The data signals are connected to the medium pins/sockets such as through pin 103.

Referring again to Figure 1, there are several signals used for the power management which are specifically shown in Figure 1. Lines 27 provide signals from the CPU blade 15 and the switch blade 14 to the blade

20 microcontrollers 20 of the MBs 10 and 12. These signals include a "presence signal" which indicates the presence of the board. Thus, when a

board is seated in the backplane, the microcontrollers on the MBs can detect the presence of each of the boards individually. Another signal is the "power good" signal. This signal indicates that a blade once installed, has the proper power. This is detected at the output of the DC-to-DC converter.

5 Another signal is a signal to indicate the health or failure of a board. Internal circuitry on the board, for instance self-test circuitry, can indicate when a blade has failed and provide a signal to the MBs indicating such failure.

The lines 26 which are the system management (SM) bus provide control signals from the MBs 10 and 12 to the other blades in the system.

10 Among the signals provided as shown at the blade 15 are a reset signal, a converter enable signal shown as "DC-DC on" and the non-maskable interrupt (NMI) signal. All of the signals discussed in connection with lines 26 and 27 may be encoded onto a serial bus such as the I²C bus for distribution to the blades.

15 In one embodiment there are two server MBs 10 and 12. One of these two blades is active and the other is in stand-by. Lines 25 in one embodiment comprising a separate bus interconnecting the MBs 10 and 12. This bus provides a signal between the MBs indicating the health of the MBs. For instance, the inactive MB may periodically ping the active MB to

20 determine its health. If the inactive blade determines that the active MB has failed it can power up and take over the system management. To this end, a

multiplexer (e.g., MUX22) is used at the output of each of the MBs to allow the selection of appropriate SM bus signals. EEPROMs on the MBs (such as EEPROM 21) allow each MB to be prepared to take over at anytime since each stores the data needed for continued operation of the system.

5 Thus, while one of the MBs handles the housekeeping chores, the other one acts as a hot spare. The peer-to-peer heartbeat monitor (or similar implementation) as mentioned is used for one blade to monitor the status of the other blade. If the active management blade fails to respond to the heartbeat "pinging," that blade is considered to fail and the hot spare

10 blade takes over. A failure LED on the face plate of the failed management blade turns on and an alert is sent out by the hot spare to warn a system administrator of the failure event.

Referring now to Figure 2A, the operation of a system is illustrated for the normal insertion of a blade. At P0 indicated by 40, the blade is separated from the backplane. At 51 (P1/H0) the blade is fully inserted into the backplane and at this time the blade provides an indication to the server MB 52 indicating its presence. This may be, for instance, the closing of a loop through the inserted blade or a signal provided by the 5V power supply. When the system MB 52 senses the presence of the blade it enables the power supply as shown by line 52. This is the signal which is connected to the terminal 112 of the DC converter as shown in Figure 5. Once the power

supply is activated, a “power good” signal may be provided to MB 54 as indicated by line 53. A “power good” signal can be provided by a power sensing chip on the blade, as an example. Thus, at H1 indicated by 55 the board is fully seated and its power is on.

5 Now the MB 54 deasserts the reset signal allowing the newly inserted blade to boot up. This is indicated by the line 57. At H2 indicated by 58, the inserted blade indicates that it is on and healthy by the health bit over the backplane or over a separate SM bus. At H3 indicated by 59, normal operation of the inserted blade occurs with the blade operating system or

10 embedded software up and running.

Figure 2B indicates a normal extraction. For normal extraction, an operator initiates a shut-down process in the operating system or a similar command in the embedded software. An alert-on-LAN message to the MB 54 occurs. The blade that will be extracted then goes through a normal shut-down and places itself into a soft-off state. This indicated in Figure 2B by the transition from H3 (60) to H2 (61) then to H1 (62). At this point in time, the blade to be extracted deasserts the health bit as indicated by line 63. Now MB 54 asserts reset and then disables the converter as indicated by line 67. Typically, an LED on the front panel of the blade to be extracted illuminates

15 indicating to the operator that the blade may now safely be removed. This is indicated by the transition from H1/H0 (70) to P0 (71).

The normal insertion and extraction is obviously the easier case to deal with from the standpoint of power management and power transients.

Figure 3 illustrates the more difficult case of a forced extraction. This case assumes that the board is extracted without first informing the MB 54 of

5 Figure 3. The transitions from 72 to 73 to 74 to 75 occur almost instantaneously. The MB 54 is able to detect this unplanned extraction of a blade, such as a CPU node, and perform the necessary remedial action.

As shown by line 76, as the blade is extracted the short pins/sockets of the connector such as pin 102 and socket 120 of Figure 5, first disengage.

10 Among the short pins/sockets that first disengage is a pin carrying the enable signal to the power supply. When this path breaks the converter is shut-down and input terminal 110 of the converter is substantially disconnected from the output terminal 111. Thus the capacitance associated with the output of the converter is disconnected from the input terminal. The 48V power supply to the input of the converter (which bypasses resistor 116) is also disengaged since it is connected to the converter on a short pin/socket.

15

At this time the “power good” signal is deasserted and the health bit is deasserted as shown by line 76. As indicated by line 77, the MB 54 deasserts the enable signal for the converter. This is done to prevent a re-powering up of the blade if the blade were to be reinserted immediately. If

this were not done and the blade was immediately reinserted, the large capacitance associated with the output of the converter would be immediately coupled through the short pin providing the direct connection to the 48V power supply. This would cause an unwanted transient which may

5 affect other blades in the system.

Next as indicated at 80, the medium length pins/sockets break contact. These include the data signals associated with the CPU and switching blades. At 81 the long pins/sockets break contact. Among the long pins that break contact at this point is the 48V power supply which is

10 coupled to the input of the converter through the resistor 116 of Figure 5. Both on extraction and on reassertion this resistor dampens the transients associated with decoupling and decoupling the converter from the backplane. This is particularly important if, for example, for some reason the power supply remains enabled during immediate reassertion. If this were to

15 occur, the first power applied to the converter would be through the resistor 116 dampening the surge which occurs when the converter is abruptly activated. By the time the shortest pins/sockets make contact this transient will have diminished. As also indicated at 81 the stand-by power being on one of the longest pins/sockets also breaks contact.

20 Figure 4 illustrates the case where a failure occurs on boot up. As indicated at 80, the blade is inserted and its presence detected by the MB

54. 81 indicates the normal activation of the power supply. 82 indicates that the blade has good power and that at this point the insertion process is proceeding normally. After H1 (83) as indicated by 84, reset is deasserted and the blade is allowed to boot up. Assume shortly thereafter, as indicated

5 at 84, failure occurs (H1F) the blade itself deactivates the power supply and informs the MB 54 of the failure. This is the response that would occur if for instance, an on-blade sensor detects an overheat condition. At this point the MB 54 asserts reset and disables the converter.

10 Thus, a power management system has been described for a system having hot swappable blades.